

WINEのUnicode化に向けて

1 Unicode化の目的

現在のWINEシステムの文字コード(内部コード)はEACC(East Asian Character Codes)だが、Millenniumの後継システムである「Sierra」はUnicode版のみしかリリースされておらず、Millenniumユーザーの大半もUnicode版を使用している。また、他の国内・海外のILS製品は基本Unicode対応のみであるという現状からみて、WINEのUnicode化への対応は急務となっている。

そのため、2015年度、本番環境とテスト環境の2台あるWINEサーバのうち、テストサーバをUnicode化し、本番サーバのUnicode化に向けて準備することとした。なお、Unicode化の例として「早」という文字は、EACCでは{214275}になり、Unicodeでは{u65E9}になる。このようにEACC-Unicode間でそれぞれ対応するコードが存在すれば、それぞれ機械的に一つずつ置き換えていく。「早稲田」の3文字は以下ようになる。

早稲田 (EACC)	→	早稲田 (Unicode)
{214275}{214f43}{214c24}		{u65E9}{u7A32}{u7530}

2 2015年度の計画

夏季一斉休業期間中にMillennium R2011. 1.4から1.6へのバージョンアップ後にテストサーバのUnicode化を実施することとした。その理由は、R2011. 1.4で生じていた問題、すなわち「巻」「札」等の文字がキーワードで検索できないなどの問題がこのバージョンアップにより修正されるといったことがある。また、本番サーバとテストサーバのバージョンを同期させたうえでテストサーバのUnicode化を行う必要があった。スケジュールは以下のとおりである。

- ・テストサーバの前日フルバックアップ取得
(2015/12/17(木) 9:00-17:00の間)
- ・テストサーバのUnicode化
(2015/12/18(金) 6:30-9:30)
- ・検証内容
 - ① 12/18の移行直後にログインができるか、一通りメニューが表示されるか、文字化けがないか確認する。なお、本格的な検証は年明けから行う。
 - ② 書誌データロード、利用者データロードについては3月末までに原則すべて検証する。ただし、メール送信等、一部テストサーバでは行えない項目がある。これらについては年明けに再度打ち合わせを行い、検討する。
 - ③ 番号取りシステムなど関連アプリは4月以降に検証する。

3 テストサーバのUnicode化実施

予定どおり12/18(金) 午前6:30からWINEテストサーバのUnicodeストレージへの移行作業を実施した。しかしながら、Innovative側キーワードインデックスの作業完了までに相当な期間を要することとなった。本番では大きな障害となるため、今後のスケジュールを綿密に立てる必要があることを認識することとなった。

4 検証結果

- ・EACCからUnicodeへ一括変換する過程でOCLC EACCとWINE EACCが異なっている文字が繁体字に変換されていることが確認できた。これらは本番サーバ移行前に修正する必要がある。
- ・キリル文字について合成文字が正しく合成されない例があった。これらについては個別にInnovativeに修正依頼を出していくこととした。
- ・また、前述のとおり、インデックス化についてInnovative社側の作業の完了が遅れたため、キーワードなどから検索できないという問題が生じた。
- ・異体字対応について香港IUGでテーブルを維持管理していることから、例えば「慶應」は「慶応」とともに検索できることを確認した。以下テストサーバの検索例。
<http://vino.wul.waseda.ac.jp/search?S12=jpn/?searchtype=X&searcharg=%E6%85%B6%E5%BF%9C&searchscope=12&SORT=DZ&extended=0&SUBMIT=%E6%A4%9C%E7%B4%A2&searchlimits=&searchorigarg=X%7Bu6176%7D%7Bu61C9%7D>
- ・以下は香港IUGのテーブル
http://hkiug.ln.edu.hk/unicode/hkiug_tsvcc_table-UnicodeVersion-1.0.html
- ・Unicodeへの変換テーブルの提供をInnovativeに要望し、EACC-Unicode変換テーブルを受領した。
- ・OCLCからのエクスポートのテストで文字化けが発生したが、InnovativeでOCLCからのロードに使用しているテーブルの仕様を修正した結果、解決した。
- ・利用者データロードの問題として、データロード前にSJISからUnicodeに変換をかける必要があることが懸念されたが、利用者ロード用テーブルの仕様の修正で対応できることがテストの結果、確認できた。

5 今後の予定

- ・今後、電子ジャーナルや電子ブックなどの書誌データのロードのテストを行い、必要に応じて使用しているロードテーブルを修正する。
- ・文字コードテーブルの修正を5月末までに検証まで終えることとした。
- ・2016年8月(一斉休業期間中)に本番サーバのUnicode化を実施する。